SUMMARY

Yellow head virus (YHV) is a major agent of disease in farmed penaeid shrimp. YHV virions purified from infected shrimp contain 3 major structural proteins of M_r 116 kDa (gp116), 64 kDa (gp64) and 20 kDa (p20). Two different staining methods indicated that the p116 and p65 proteins are glycosylated. Here we report the complete nucleotide sequence of the ORF3 which encodes a polypeptide of 1,664 amino acids with calculated M_r of 185,713 kDa (pI = 6.68). Hydropathy analysis of the deduced ORF3 protein sequence identified 6 transmembrane helices and 3 ectodomains containing multiple sites for N-linked and O-linked glycosylation. N-terminal sequence analysis of mature gp116 and gp64 proteins indicated that each was derived from ORF3 by proteolytic cleavage of the polyprotein between residues Ala^{228} and Thr^{229}, and Ala^{1127} and Leu^{1128} located at the C-terminal side of transmembrane helices 3 and 5, respectively. Comparison with the deduced ORF3 protein sequences of Australian gill-associated virus (GAV) indicated 83% identity in gp64 and 71% identity in gp116 which featured 2 significant sequence deletions near the N-terminus. Database searches revealed no significant homology with other proteins. The recombinant gp64 with and without the C-terminal transmembrane region expressed in *E. coli* was shown to react with antibody raised against native gp64 purified from virions.

Introduction

Viral disease is a serious problem for the large and rapidly expanding shrimp culture industry in Asia and the Americas, causing mass mortalities in ponds and heavy production losses [Flegel, 1998]. In the Asian region, yellow head disease has been a major concern since it first emerged in Thailand in 1990 (Limsuwan, 1991). Gross signs often associated with yellow head disease include cessation of feeding, swimming near the surface and pond edges and the development of yellow coloration of the cephalothorax and gills [Chantanacookin *et al.*, 1993]. Lymphoid organs of moribund shrimp display evidence of necrosis and contain vacuolated cells with hypertropied nuclei and densely basophilic cytoplasmic inclusions (Chantanachookin *et al.*, 1993). Death of infected shrimp usually occurs within 2-3 days of the first appearance of visible signs of disease.

Yellow head virus (YHV) is an enveloped, rod-shaped particle (approximately 40 nm x 170 nm) with prominent surface projections (approximately 11 nm) and an inner helical nucleocapsid (Chantanachookin *et al.*, 1993; Wang & Chang 2000; Loh *et al.*, 1997). Primarily based on the virion morphology and the presence of a single-stranded RNA genome (Wongteerasupaya *et al.*, 1995), YHV was previously reported as a rhabdovirus (Nadala *et al.*, 1997). However, it was subsequently demonstrated that the YHV genome is positive-sense RNA [Tang & Lightner, 1999]. Sequence analysis has also revealed that, like closely related gill-associated virus (GAV) from Australia (Cowley *et al.*, 1999; Cowley et al., 2000), YHV contains a large replicase gene (ORF1b) that is expressed as a polyprotein by ribosomal frame-shift at a "slippery" sequence that follows a predicted pseudoknot structure (Sittidilokratna *et al.*, 2002). Considerations of sequence identity, genome organization and gene expression have indicated that GAV and YHV are related to coronaviruses, toroviruses and arteriviruses and will be classified in new taxa (proposed family *Roniviridae*, genus *Okavirus*) within the order *Nidovirales* (Cowley *et al.*, 2000; Sittidilokratna *et al.*, 2002).

Nadala *et al.*, 1997 originally reported that YHV particles comprise 4 structural proteins of approximately 170 kDa, 135 kDa, 67 kDa and 22 kDa, of which the 135 kDa protein was glycosylated. However, Wang and Chang (2000) subsequently reported only 3 major YHV proteins (110 kDa, 63 kDa and 20 kDa), suggesting the larger protein may be of cellular origin. In GAV, structural proteins are encoded in ORF2 and ORF3 located immediately downstream of the ORF1b gene (Cowley *et al.*, 2001). ORF2 encodes a 22

kDa structural protein that functions as the nucleoprotein (Cowley *et al.*, submitted), and ORF3 encodes a polypeptide with the structural characteristics of a large glycoprotein with multiple membrane-spanning domains (Cowley and Walker, 2002). In this paper, we report the nucleotide and deduced amino acid sequences of the YHV ORF3 region. We show that the region encodes the major viral structural glycoproteins (gp116 and gp64) that are synthesized as a polyprotein which undergoes post-translational modification including proteolytic cleavage and glycosylation.

METHODS

Viral purification

As a source of material for virus purification, 200-500 juvenile *Penaeus monodon* shrimp (average weight 20 g), were infected by intramuscular injection with a standard inoculum of a 1998 Thai YHV isolate prepared as described previously (Wongteerasupaya *et al.*, 1995; Sittidilokratna *et al.*, 2002). At 3 days post-infection, haemolymph was withdrawn and used as a source of virus for purification. Virions were purified by Urograffin (Schering AG, Germany) gradient ultracentrifugation as described by Wongterasupaya *et al.* (1995). The stocks of purified virus were snap-frozen and stored at -80°C until required for analysis.

SDS-PAGE and N-terminal sequencing

Protein samples (10 µg) were analyzed by SDS-PAGE (Laemmli, 1970) under reducing conditions in 7.5% or 15% discontinuous gels. The fractionated proteins were transferred to a polyvinylidene difluoride (PVDF) membrane in transfer buffer (10 mM CAPS 10% methanol) using a semi-dry blotter (Hoeffer). Protein bands blotted on the membrane were stained briefly in 0.3% Coomassie brilliant blue R250. The protein bands were then excised and analyzed by N-terminal sequencing. Edman degradation was conducted using an Applied Biosystems Sequencer at the Center for Genetic Engineering and Biotechnology, Bangkok.

Polyclonal antibody production and immunoblot analysis

Nitrocellulose-bound antigen was prepared as described by Diano *et al.* (1998). Briefly, 500 µg of total viral protein were blotted onto a nitrocellulose membrane (Amersham Pharmacia, Upjohn), excised and ground with 0.5 ml phosphate buffer saline (PBS) [140mM NaCl, 4mM KCl, 2mM KH_2PO_4, 8mM Na_2HPO_4, pH 7.4] in liquid nitrogen. This antigen was mixed with an equal volume of complete Freund's adjuvant (Sigma) and used to immunize two BALB/C mice by an intraperitoneal injection. Mice were boosted at one week intervals with 100 µg of viral protein in incomplete Freunds adjuvant. The antisera were collected one week after the second boost and used for immunoblot analysis.

Proteins were transferred to a nitocellulose membrane (Amersham Pharmacia, Upjohn) using a semi-dry blotting apparatus (Hoeffer). Following electroblot tranfer, the membrane was blocked in 3% bovine serum albumin (BSA), 0.5% Tween 20 in PBS for 3 h. The membrane was washed briefly in the same buffer without BSA and then reacted with 1:10,000 dilution of polyclonal antibodies for 1 h. Goat anti-mouse polyclonal antibodies conjugated with alkaline phosphatase were then reacted for 2 h. Immuno-reactive bands on the membrane were visualized by adding nitro blue tetrazolium (NBT) and 5-bromo-4-chloro-3-indoyl phosphate (BCIP).

Glycoprotein staining

Glycoprotein detection was performed using the ECL glycoprotein detection system (Amersham Pharmacia). Proteins blotted on PVDF membranes were oxidized with 10 mM sodium metaperiodate in 100 mM acetate buffer pH 5.5 at room temperature for 20 min in the dark. The membrane was then reacted with 0.35 mM biotin hydrazide in 100 mM acetate buffer pH 5.5 for 1 h followed by incubation with streptavidin conjugated with alkaline phosphatase. The glycoprotein bands were visualized by adding NBT and BCIP. The thymol staining was essential the same as described by Racusen *et al.* (1979).

Reverse transcriptase polymerase chain reaction (RT-PCR), cloning and sequencing

YHV genomic RNA was extracted from purified virus using Trizol Reagent (Invitrogen, U.S.A.), resuspended in DEPC-treated water. The RT-PCR was performed in a total volume of 25 µl containing 200 ng of genomic RNA, 0.4 mM of forward primer (5`-gatcggggtacctaagcttatgctatcgaccta-3`) designed from the 3`-end of the ORF1b gene and oligo d(T) primer (5`-tctagaggatccccgggtaccttttttttttttttttttt-3`) using SuperScript one-step RT-PCR system (Invitrogen, U.S.A.) in the presence of 1 unit of Elongase (Invitrogen, U.S.A.) according to the instruction manual. The RT-PCR profile consisted of an initial incubation at 50°C for 30 min, 94°C for 2 min followed by 35 cycles of amplification. Each cycle consisted of denaturation at 94°C for 30 sec, annealing/extension at 68°C for 8 min. The RT-PCR product was either sequenced directly or cloned in pGEM-T Easy vector. The RT-PCR product was gel-purified using QIAquick Gel Extraction kit (Qiagen) and directly sequenced using BIG Dye™ reagent (ABI, U.S.A.). Nucleotide sequences obtained from an initial sequencing were used to

design new primers to sequence toward the 5`-end and the 3`-end of the fragment. Overlapping sequences were then analysed using SeqEd 1.0.3 (ABI, U.S.A.).

Recombinant protein expression in *E. coli*

The cDNA encoding full length gp64 was generated by PCR using forward primer (YHV-D7) 5`-GCCTCTAGACATATGCTCGCTCCACGACAGGCACGTGTT-3` and reverse primer (YHV-D8)

5`-CATTGTGGATCC**TCACTA**GTGATGATGATGATGATGGGATCGTTTGGCTTT-

CGTTCTCATGGACGT-3`. The forward primer was designed from residues L^{1128} to V^{1135} in the ORF3 polyprotein and included initiation Met and an *Nde*I restriction site (underline) while the reverse primer was designed from residues T^{1657} to S^{1666} and included stop codons (bold) and a *Bam*HI restriction site (underline). PCR was performed in a 25 µl reaction mixture containing 1X PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X-100), 0.2 mM of each dNTP, 1 ng oligo-primed cDNA, 0.25 µM of each primer and 2 units of *Taq* DNA polymerase (Perkin Elmer, U.S.A.). The reaction mixture was subjected to 35 cycles of denaturation at 94°C 30 sec, annealing/extension at 68°C 2 min, and followed by the final extension at 72°C for 10 min. The PCR products were cut with *Nde*I and *Eco*RI and ligated into the multiple cloning site of pET17b (Novagen). The C-terminal transmembrane deleted construct was generated to remove the last 40 C-terminal residues (Y^{1627}-S^{1666}) by PCR using YHV-D7 and YHV-D14 [5`-GAATTC**TCACTA**ATCCCATGTCTTGCCGCCGAA; corresponds to residues $F^{1620}$-$D^{1626}$ with an *Eco*RI attached at the 5`-end and stop codons (bold)]. The PCR product was cloned into pDrive vector (Qiagen) and sequenced. The insert was then digested with *Nde*I and *Eco*RI and cloned into *Nde*I/*Eco*RI sites of pET17b. The recombiant plasmids were then transformed into *E. coli* BL21(DE3) (Novagen). Overnight culture of BL21(DE3) was 1:20 diluted with fresh LB broth and grown at 37°C to an A_{600} of 0.5-0.6. The culture was then induced with 1 mM IPTG and induced for 6 h. One milliliter of culture was pelleted, suspended in protein loading buffer and heated at 100°C for 10 min before subjected to SDS-PAGE. The recombinant protein produced in *E. coli* was confirmed by immunoblot analysis using gp64 polyclonal antiserum.

RESULTS

Analysis of structural proteins of YHV virions

Transmission electron microscopy of YHV virions purified from shrimp hemolymph revealed the typical rod-shaped, enveloped particles described by Wongteerasupaya *et al.* (1995) [data not shown]. The purified particles were disrupted with SDS and analyzed by SDS-PAGE. As shown in Figure 1, analysis under reducing conditions revealed three relatively abundant proteins of M_r 116 kDa, 64 kDa, and 20 kDa. Similar results were obtained by gel electrophoresis under non-reducing conditions (data not shown), indicating that the virion proteins were not linked by intermolecular disulfide bonds. As shown in Figure 1, hemocyanin (~68-70 kDa) is the most abundant protein in shrimp hemolymph (Figueroa-Soto *et al.*, 1997). As there was little evidence residual hemocyanin following purification, the virus preparation appeared to be largely free of cellular proteins.

The YHV structural proteins were examined for evidence of glycosylation using two different detection methods. The ECL glycoprotein detection assay is based on oxidation of the ketone group of sugar residues with sodium metaperiodate and conjugation of the oxidized ketone ring with biotin [Murray *et al.*, 1989]. The carbohydrate can then be visualized *in situ* on a membrane. The second method is used the thymol reagent and is based on the hydrolysis of glycosidic bonds concomitant with the formation of furfural derivative (Racusen, 1979). As shown in Figure 1B and 1C, each method indicated that the 116 kDa and 64 kDa proteins were glycosylated. There was no evidence of glycosylation of the 20 kDa protein.

Each of the three YHV structural proteins (gp116, gp64 and p20) was subjected to N-terminal sequence analysis. The N-terminal sequences for glycoproteins gp116 and gp64 were determined to be T-I-L-S-G-I-P-E-K-D- and L-A-P-R-Q-A-R-V-X-G- (X, uncertain residue), respectively. No sequence was obtained for protein p20 which appeared to be blocked at the N-terminus.

Nucleotide sequence and deduced amino acid sequence of ORF3

A region of the YHV genome extending from the 3'-poly[A] tail to a locus at the 3'-end of the ORF1b gene was amplified by RT-PCR. Analysis of the amplified product

by agarose gel electrophoresis revealed a single band of approximately 6.0 kbp. By comparison with the known complete sequence of GAV (Cowley *et al.*, 2000; Cowley & Walker, 2002) and available partial sequence data on the YHV genome (Sittidilokratna and associates, unpublished data), the size of the amplified product was consistent with the expected size of 3'-end of YHV genome (Figure 2). Nucleotide sequence analysis revealed the region contained 2 long open reading frames in the same sense (+) as ORF1b. The largest open reading frame (ORF3) commenced 848 nucleotides downstream of the ORF1b termination codon and comprised 4998 nucleotides. Analysis of the nucleotide 240 bp upstream of the initiation codon using Sigscan (Heinemeyer et al., 1998) and TF-SEARCH (Prestridge, 1991) databases showed no typical TATA box but contained one inverted CCAAT box at positon -49/-37. However there is a TATA-like sequence (TAAAT) found at postions -29 -24. The complete nucleotide sequence and deduced amino acid sequence of YHV ORF3 is shown in Fig. 3. ORF3 encodes a polypeptide of 1,666 amino acids with predicted molecular weight of 185,713 and pI of 6.68. Alignment with the N-terminal sequences of mature gp116 and gp64 identified perfect identity with residues T^{229} to F^{236} and L^{1128}-V^{1135} respectively of the encoded polypeptide. The data indicated that gp116 and gp64 are derived by post-translational proteolysis of the ORF3 polyprotein. Based on the identified N-terminal sequences and presumed sites of proteolysis, the calculated molecular weight for unprocessed gp116 was 101,734 Da and for gp64 was 58,599 Da. The predicted size of these products was consistent with evidence of post-translational glycosylation of each protein (Fig. 1B and 1C).

Hydropathy plot of the ORF3 polypeptide sequence (http://sosui.proteome.bio.tuat.ac.jp) identified six hydrophobic transmembrane helices - three in the N-terminal domain (residues L^{25}-L^{47}, Y^{138}-F^{155}, F^{208}-I^{230}), two in the central domain (residues G^{1027}-N^{1049} and F^{1100}-L^{1128}) and one in the C-terminal domain (residues K^{1630}-L^{1652}). Analysis of membrane topology using TMHMM (Krogh *et al.*, 2001) predicted that the N-terminus of the ORF3 polyprotein is anchored inside the cell via the first transmembrane domain. Three subsequent ectodomains are connected by the second to the sixth transmembrane domains respectively, and the C-terminus of the protein is also anchored within the cell (Fig. 4). Proteolytic cleavage between A^{228} and T^{229}, and A^{1127} and L^{1128} near C-terminus of transmembrane helices 3 and 5 respectively would generate gp116 and gp64. The model predicts that gp64 is a type I transmembrane glycoprotein anchored in the viral envelope at the C-terminus (Fig. 4). According to the predicted

topology, gp116 is a polytopic type III transmembrane glycoprotein, anchored at the C-terminus but with both C-terminus and N-terminus external. The model also predicts the formation of an unidentified 25.4 kDa protein with three membrane-spanning domains and type III membrane topology. Prosite analysis using NetOGlyc2.0 database (http://www.cbs.dtu.dk) [Hansen *et al.*, 1997] revealed the presence of *O*-linked glycosylation sites at threonine residues 255, 414, 569, 807, 828, 1585 and at serine residues 70, 232, 413, 417, 576 and 832 of the ectodomains of the ORF3 polyprotein. As shown in Fig. 3, potential N-linked glycosylation sites were also identified in the ectodomains of gp64 (5 sites), gp116 (7 sites) and the unidentified 25.4 kDa transmembrane fragment of ORF3 (1 site). The ectodomains were also rich in cysteine residues, suggesting complex folded loop structures in the mature glycoproteins (Fig. 3).

FASTA, TFASTA and BLAST searches of the deduced gp116 and gp64 sequences against GenBank/EMBL, SWISS-PROT and PIR protein databases revealed no significant similarity with other proteins including the spike glycoproteins of other nidoviruses (i.e. coronaviruses, toroviruses and arteriviruses). However, comparison with the available deduced sequence of the GAV ORF3 polyprotein (Cowley and Walker, 2002) indicated overall amino acid sequence similarity of 75%. The level of sequence homology was highest in gp64 (83% similarity) and lowest in gp116 (71% similarity), primarily due to 2 significant deletions at the N-terminus of the gp116 homologue in GAV (see Fig. 5). A comparison of individual sites in the YHV and GAV ORF3 polyproteins indicated that all cysteine residues were conserved and all but two (N^{291}-S^{293} and N^{711}-T^{713} in gp116) of the potential YHV N-glycosylation sites were preserved. Three of eight predicted O-glycosylation sites in YHV gp116 were not conserved and the single O-glycosylation site in YHV gp64 was not conserved in GAV, suggesting this protein may contain only N-glycans.

Expression of gp64 in *E. coli*

The gp64 coding region of YHV ORF3 was amplified by PCR using a forward primer designed from L^{1128} to V^{1135} in ORF3 with an upstream initiation codon and a reverse primer designed from T^{1657} to S^{1666} including the endogenous translation termination codon. The PCR product was sequenced and cloned into pET17b expression vector. Recombinant plasmid pET17b-gp64 was transformed into *E. coli* BL21(DE3). As shown in Fig. 6B, expression of recombinant gp64 was not evident by SDS-PAGE

analysis of the *E. coli* whole cell lysate. However, by immunoblot analysis using polyclonal antibody against the native gp64 purified from virions, an immunoreactive ~ 58 kDa band was detected on blot (lanes 3-6). The polyclonal antibody also reacted with native gp64 in purified virus (lane 7) but did not react with lysates prepared from *E. coli* transformed with pET17b alone (lane 1). Although the expression level is relatively poor, the size of the band is consistent with the molecular weight of gp64 calculated from the deduced amino acid sequence. The difference is M_r between the native and recombinant proteins is most likely due to the absence of post-translational glycosylation in *E. coli*. We also observed that the level of ~58 kDa band decreased 1 h after induction, indicating the recombinant gp64 produced in *E. coli* is not stable . It is possible that the transmembrane region of gp64 may cause poor expression. We then generated the transmembrane deleted construct(pET16b-gp64ΔTM, see Fig. 6A) by removing residues Y^{1627}-S^{1666} located at the C-terminal of gp64. This construct was transformed in *E. coli* and induced with IPTG. Again there was no evidence of overexpressed protein band corresponds to the size of gp64 lacking transmembrane domain (~55 kDa) on Coomassie stained gel. However Western analysis of this gel revealed the presence of strong immunoreactive band approximately 55 kDa. This band was increased over the induction period (1-5 h) suggesting that recombinant gp64 without transmembrane domain is more stable than full length gp64.

DISCUSSION

In mammalian nidoviruses, the virion envelope glycoproteins have functions that are principally associated with recognition of target cells, the fusion of viral and cellular membranes and the release of virions from infected cells. The envelope glycoproteins also have a central role in the induction of the host immune response and are targets for protective immunity (for a review see Spaan *et al.*, 1988). In some vertebrate nidoviruses, the surface glycoproteins may also play an important role in the determination of virulence (Dalziel *et al.*, 1987; Phillips *et al.*, 1999). Little is presently known of the structure or function of the envelope glycoproteins of the newly discovered invertebrate nidoviruses infecting shrimp or of the defensive response of shrimp (or any other invertebrate) to viral infection. In this paper, we report an initial characterization of invertebrate nidovirus glycoproteins. We show that YHV virions contain three major structural proteins including a ~20 kDa non-glycosylated protein and two envelope glycoproteins (gp116 and gp64) that are expressed as a polyprotein from ORF3 and processed by post-translational proteolysis. We also show that recombinant gp64 with or without transmembrane region could be expressed in *E. coli* in a form that is recognized by antibody to the native virion glycoprotein.

Nadala *et al.* 1997) have previously reported that YHV contains 4 major structural proteins (Mr = 200 kDa, 135 kDa, 62 kDa and 20 kDa) of which only the 135 kDa protein was reported to be glycosylated. Wang & Chang (2000) have reported only three structural proteins (Mr = 110 kDa, 63 kDa and 20 kDa). Our results are in close agreement with this more recent report and we demonstrate that each of the larger structural proteins is glycosylated. Small differences in the M_r of the structural proteins could well be due to differences in the conditions of electrophoresis or to variations in the pattern of glycosylation of different YHV isolates.

N-terminal sequence analysis of gp116 and gp64 allowed precise identification of sequences encoding these proteins within the 1,666 amino acid ORF3 polyprotein. The size of the YHV polyprotein is similar to those of vertebrates coronaviruses including avian infectious bronchitis virus (1,160 amino acids) [Binns *et al.*, 1985], feline infectious peritonitis virus (1,452 amino acids) [de Groot *et al.*, 1987], the murine hepatitis virus (1,376 amino acids) [Spann *et al.*, 1988], porcine epidemic diarrhea virus (Duarte & Laude, 1994) [1,383 amino acids] and the human respiratory coronavirus (1,353 amino

acids) [Mounir & Talbot, 1993]. However, the predicted structure of the ORF3 polyprotein is highly complex, comprising 6 transmembrane regions, three ectodomains and two intra-virion domains. By contrast, coronaviruses contain a single transmembrane region located near the C-terminus of the protein.

The envelope spikes of vertebrate coronaviruses and toroviruses comprise a large (~180 kDa) glycoprotein (S) that is often cleaved by a cellular protease to yield two similarly sized subunits (S1 and S2) which remain non-covalently associated in virions (Sturman *et al.*, 1985; Cavanagh, 1995; Snijder & Horzinek, 1995). In YHV, gp64 and gp116 are also generated by proteolysis of a ~180 kDa polyprotein. Although each is rich in cysteine residues, SDS-PAGE under non-reducing conditions indicated that gp116 and gp64 are not covalently associated. It is possible that, like the S1 and S2 subunits of coronavirus surface glycoprotein, gp64 and gp116 are associated non-covalently to form the prominent peplomers on the virion surface. In coronaviruses, the S2 subunit appears to form the membrane-bound stalk while the S1 subunit forms the globular head of the spike (de Groot *et al.*, 1987) which interacts with the host cell receptor (Kubo *et al.*, 1994; Suzuki & Taguchi, 1996). The S2 subunit contains several functional domains including a membrane anchor, the six strictly conserved cysteine residues and a leucine zipper motif (Britton, 1991) which has been shown to mediate the oligomerization of the subunits and induce cell fusion (Grosse & Siddell, 1994; Luo *et al.*, 1999). The S2 ectodomain also contains two large amphiphatic α-helices with a heptad repeat that have been proposed to mediate coiled-coil interchain interactions (de Groot *et al.*, 1987). As reported for GAV (Cowley & Walker, 2002), examination of YHV gp64 and gp116 sequences revealed no heptad repeats or significant amphiphatic α-helices suggesting an absence of coiled-coil structures. Further work is required to determine if the processed gp116 and gp64 are associated to form the envelope spikes.

Sequence analysis of the of the YHV genome was facilitated by use of an oligo (dT) primer to amplify the 3'-terminal region. This approach was adopted in the expectation that the YHV genome would be polyadenylated - a characteristic of the viruses (+) RNA viruses in the order *Nidovirales*. The ORF3 gene, encoding the YHV structural glycoproteins, is located downstream of the ORF1b gene which encodes replication enzymes and has features characteristic of nidoviruses (Sithidilokratana *et al.*, 2002). In GAV, ORF2 (located in the genome between ORF1b and ORF3) appears to

encode the nucleoprotein (Cowley *et al.*, submitted) and this corresponds to the major YHV virion protein p20. Clearly, GAV and YHV are closely related in sequence and share a common gene organization which appears unique amongst nidoviruses in that the nucleoprotein gene is upstream of the structural glycoprotein gene and there is no discrete gene encoding the integral membrane (M) protein. However, it is evident from the location of proteolytic cleavage sites in the YHV ORF3 polyprotein that the N-terminal fragment has not yet been identified in virions or infected cells. If no further cleavage occurs, this 227 amino acid (25.4 kDa) fragment will be of similar size to the integral membrane proteins of coronaviruses and toroviruses (225-262 amino acids) which occur abundantly in virions and appear to have a role in intracellular budding (Rottier, 1995). The YHV ORF3 fragment also contains 3 membrane-spanning domains characteristic of the M proteins but the predicted membrane topology is in the reverse orientation with an N-terminal cytoplasmic tail and the C-terminus oriented towards the virion surface. Also, as p20 appears to be the nucleoprotein encoded in ORF2 (Cowley *et al.*, submitted), there is no evidence that YHV has a major virion component corresponding to the coronavirus M protein. The use of antibodies against suitable peptides in the N-terminal domain should assist identification and characterization of the remaining fragment of the ORF3 polyprotein.

The molecular weights of gp116 and gp64 calculated from the nucleotide sequence are smaller (~15 kDa and ~4 kDa respectively) than the estimated size of the mature forms as determined by SDS-PAGE. The size differences are evidently due to glycosylation as revealed by the detection of carbohydrate (Fig. 1) and the presence of several putative N- and O-linked glycosylation sites in each protein. The authenticity of cDNA encoding gp64 was also confirmed by the expression in *E. coli*. Although the expression level of recombinant gp64 was relatively poor, it was readily detected using polyclonal antibodies raised against the native form. The size of recombinant gp64 was consistent with the molecular weight calculated from deduced amino acid sequence. Poor expression of gp64 in *E. coli* suggests that the protein is unstable and may not be properly folded when expressed in *E. coli* independently of gp116. Our attempt to express gp64 as the fusion protein with glutathione S-transferase failed to improve the yield. The expression of gp64 in *E. coli* has been improved only after removal of transmembrane domains (Fig. 6C). Similar result was observed when gp116 with and without transmembrane domain were expressed in *E. coli* (data not shown). Expression of these proteins in insect cells which

may provide an intracellular environment similar to shrimp cells, providing a more useful approach to investigations of the processing and assembly of the glycoproteins encoded in YHV ORF3.

Overall, these data indicate that, despite defining similarities in the ORF1b region, YHV has a genome organization and glycoprotein expression strategy that is fundamentally different from other nidoviruses. The work further supports the classification of YHV with GAV in the new family *Roniviridae* (Cowley *et al.*, submitted; Sithidilokratana *et al.*, 2002).

Acknowledgements

This work was supported by the National Center for Genetic Engineering and Biotechnology (BIOTEC), the Thailand Research Fund (TRF) and the Australian Centre for International Agricultural Research (ACIAR). S.J. is a recipient of the TRF postdoctoral fellowship. S.U. is supported by Graduate Studies Scholarship, Mahidol University and the TRF senior research scholarship, N.S. is supported by Graduate Studies Scholarship, Mahidol University and the Ph.D. Golden Jubilee program, TRF. S.P. is a TRF senior research scholar.

References

1. Binns, M.M., Boursnell, M.E.G., Cavanagh, D., Pappin, D.J.C. and Brown (1985) Cloning and sequencing of the gene encoding the spike protein of the coronavirus IBV. *J. Gen. Virol.* **66**, 719-72
2. Britton, P. (1991) Coronavirus motif. *Nature* **353**, 394.
3. Cavanagh, D. (1995) The coronavirus surface glycoprotein. In: The Coronaviridae. S.G. Siddle (ed) p73-113. Plenum Press, New York
4. Chantanachookin, C., Boonyaratpalin, S., Kasornchandra, J., Sataporn, D., Ekpanithanpong, U., Supamataya, K., Sriurairatana, S. and Flegel, T.W. (1993) Histology of ultrastructure reveal a new granulosis-like virus in *Penaeus monodon* affected by yellow-head disease. *Diseases of Aquatic Organisms* **17**, 145-157.
5. Cowley, J.A., Cadogan, L.C., Spann, K.M. and Walker, P.J. (2002) The ORF2 gene encodes the nucleocapsid protein of gill-associated nidovirus of *Penaeus monodon* prawns. *J. Gen. Virol.* submitted
6. Cowley, J.A., Dimmock, C.M., Spann, K.M. and Walker, P.J. (2000) Detection of Australian gill-associated virus (GAV) and lymphoid organ virus (LOV) of *Penaeus monodon* by RT-nested PCR. *Disease of Aquatic Organisms* **39**, 159-167.
7. Cowley, J.A., Dimmock, C.M., Spann, K.M. and Walker, P.J. (2001) Gill-associated virus of *Penaeus monodon* prawns: Molecular evidence for the first invertebrate nidovirus. In: The Nidoviruses. E. Lavi, S.R Weiss and S.T. Hingley (eds.). Advances in Experimental Medicine and Biology 494, 43-48. Kluwer Academic/Plenum Press, New York.
8. Cowley, J.A. & Walker, P.J. (2002) The complete sequence of gill-associated virus of *Penaeus monodon* prawns indicates a gene organisation unique among nidoviruses. *Arch. Virol (In Press).*
9. Dalziel, R.G., Lampert, P.W., Talbot, P.J. and Buchmeier, M.J. (1986). Site-specific alteration of murine hepatitis virus type 4 peplomer glycoprotein E2 results in reduced neurovirulence. *J. Virol.* **59**, 463-471.
10. de Groot, R.J., Maduro, J., Lenstra, J.A., Horzinek, M.C., van Der Zeijst, B.A.M. and Spaan, W.J.M. (1987) cDNA cloning and sequence analysis of the gene encoding the peplomer protein of feline infectious peritonitis virus. *J. Gen. Virol.* **68**, 2639-2646.
11. De Groot, R.J., , Luytjes, W., Horzinek, M.C., van der Zeijst, B.A.M., Spann, W.J.M. and Lenstra, J.A. (1987) Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *J. Mol. Biol.* **196**, 963-966.

12. Diano, M., Le Bivic, A. and Hirn, M. (1998) Raising polyclonal antibodies using nitrocellulose-bound antigen. *Methods. Mol. Biol.* **80**, 5-13.

13. Duarte, M. and Laude, H. (1994) Sequence of the spike protein of the porcine epidemic diarrhoea virus. *J. Gen. Virol.* **75**, 1195-1200.

14. Figueroa-Soto, C.G., de la Barca, A.M.C., Vazquez-Moreno, L., Higuera-Ciapara, I. and Yepiz-Plaascencia, G. (1997) Purification of hemocyanin from white shrimp (*Penaeus vannamei* Boone) by immobilized metal affinity chromatography. *Comp. Biochem. Physiol.* **117B**, 203-208.

15. Flegel, T.W. (1997) Major viral diseases of the black tiger prawn (*Penaeus monodon*) in Thailand. *World J. Microbiol. Biotechnol.* **13**, 433-442.

18. Grosse, B. and Siddell, S.G. (1994) Single amino acid changes in the S2 subunit of the MHV surface glycoprotein confer resistance to neutralization by S1-specific monoclonal antibody. *Virology* **202**, 814-824.

19. Hanen, J.E., Lund, O., Rapacki, K. and Brunak, S. (1997) O-glycbase version 2.0-A revised database of O-glycosylated proteins. *Nucleic Acids Res.* **25**, 278-282.

20. Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A.E., Kel, O.V., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Kolpakov, F.A., Podkolodny, N.L. and Kolchanov, N.A. (1998) *Nucleic Acid Res.* **26**, 362-367.

21. Krogh, A., Laarsson, B., von Heijne, G. and Sonnhammer, E.L.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567-580.

22. Kubo, H., Yamada, Y.K. and Taguchi, F. (1994) Localization of neutralization epitopes and the receptor-binding site within the amino-terminal 330 amino acids of the murine coronavirus spike protein. *J. Virol.* **68**, 5403-5410.

23. Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

24. Limsuwan, C. (1991) Handbook for cultivation of black tiger prawns. Tansetakit Co. Ltd., Bangkok (in Thai).

25. Loh, P.C., Tapay, L.M., Lu, Y. and Nadala, E.C. Jr. (1997) *Adv. Virus Res.* **48**, 263-312.

26. Luo, Z., Matthews, A.M. and Weiss, S.R. (1999) Amino acid substitutions within the leucine zipper domain of the murine coronavirus spike protein cause defects in oligomerization and the ability to induce cell-to-cell fusion. *J. Virol.* **73**, 8152-8159.

27. Mounir, S. and Talbot, P. (1993) Molecular characterization of the S protein gene of

human coronavirus OC43. *J. Gen. Virol.* **74**, 1981-1987.

28. Murray, M.C., Bhavanadan, V.P. and Davidson, E.E. (1989) *Carbohy. Res.* **186**, 255-265.
29. Nadala, E.C.B., Tappy, L. M. And Loh, P. C. (1997) Yellow-head virus: a rhabdovirus-like pathogen of penaeid shrimp. *Diseases of Aquatic Organisms* **31**, 141-146.
30. Phillips, J.J., Chua, M.M., Lavi, E. and Weiss S.R. (1999). Pathogenesis of MHV4/MHV-A59 recombinant viruses: the murine coronavirus spike protein is a major determinant of neurovirulence. *J. Virol.* **73**, 7752-7760.
31. Prestridge, D.S. (1991) *CABIOS* 7, 203-206.
32. Racusen, D. (1979) Glycoprotein detection in polyacrylamide gel with thymol and sulfuric acid. *Anal. Biochem.* **99**, 474-476.
33. Rottier, P.J.M. (1995). The coronavirus membrane glycoprotein. . *In*: The Coronaviridae. S.G. Siddell (ed.) Pp. 115-139. Plenum Press, New York.
34. Sittidilokratana, N., Hodgson, R.A.J., Panyim, S., Cowley, J.A., Jitrapakdee, S., Boonsaeng, V. and Walker, P.J. (2002) The complete *ORF1b*-gene sequence indicates yellow head virus is an invertebrate nidovirus. *Dis. Aquatic Organisms* **50**, 87-93.
35. Snijder, E.J. and Horzinek, M.C. (1995). The molecular biology of toroviruses. *In*: The Coronaviridae. S.G. Siddell (ed.) Pp. 219-238. Plenum Press, New York.
36. Spann, K.M., Vickers, J.E. and Lester, R.J.G. (1995) Lymphoid organ virus of *Penaeus monodon* from Australia. *Diseases of Aquatic Organisms* 23, 127-134.
37. Spann, K.M., Cowley, J.A., Walker, P.J. and Lester, R.J.G. (1997) Gill-associated virus (GAV), a yellow head-like virus from *Penaeus monodon* cultured in Australia. *Diseases of Aquatic Organisms* **31**, 169-179.
38. Spaan, W., Cavanagh, D. and Horzinek, M.C. (1988) Coronavirus structure and genome expression. *J. Gen. Virol.* **69**, 2939-2952.
39. Stephens, E.B. and Compans, R.W. (1988) Assembly of animal viruses at cellular membranes. *Annu. Rev. Microbiol.* **42**, 489-516.
40. Sturman, L.S., Ricard, C.S. and Holmes, K.V. (1985) Proteolytic cleavage of the E2 glycoprotein of murine coronavirus: activation of cell-fusing activity of virions by trypsin and separation of two different 90K cleavage fragments. *J. Virol.* **56**, 904-911.
41. Suzuki, H. and Taguchi, F. (1996) Analysis of the receptor binding site of murine coronavirus spike glycoprotein. *J. Virol.* **70**, 2632-2636.

42. Thompson, J.D. Higgins, D.G. and Gibson, T.L. (1994) *Nucleic Acid Res.* **22**, 4673-4680.

43. Tang, K. F.-J. and Lightner, D.V. (1998) A yellow head virus probe: application to in situ hybridization and determination of its nucleotide sequence. *Diseases of Aquatic Organisms* **35**, 165-173.

44. Wang, Y.-C. and Chang, P.-S. (2000) Yellow head virus infection in the giant tiger prawn *Penaeus monodon* cultured in Taiwan. *Fish Pathology* **35**, 1-10.

45. Wongteerasupaya, C., Sriurairatana, S., Vicker, J.E., Akrajamorn,S., Boonsaeng,V., Panyim, S., Tassanakajon, A., Withyachumnarnkul, B. And Flegel, T.W. (1995) Yellow-head virus of *Penaeus monodon* is an RNA virus. *Diseases of Aquatic Organisms* **22**, 45-50.

Figure legends

Figure 1. Analysis of structural proteins of YHV. Purified YHV was subjected to 15% discontinuous SDS-PAGE and stained with Coomassie brilliant blue (A) or transferred to nitrocellulose and stained using ECL glycoprotein detection system (B) or thymol (C). M, molecular weight protein markers; 1, purified YHV; 2, hemocyanin from shrimp hemolymph. Arrows indicate three distinct sizes of YHV proteins.

Figure 2. Schematic diagram of the organization of GAV and YHV genomes. Boxes represent the open reading frames (ORF) of subgenomic RNA of each virus spanning over 25 kb of genomic RNA. The discontinuous boxes between ORF1a and ORF1b represent the overlapped (-1) ribosomal frameshift site (Cowley *et al.*, 2000; Sithidilokratana *et al.*, 2002). N, nucleocapsid; S protein, structural proteins.

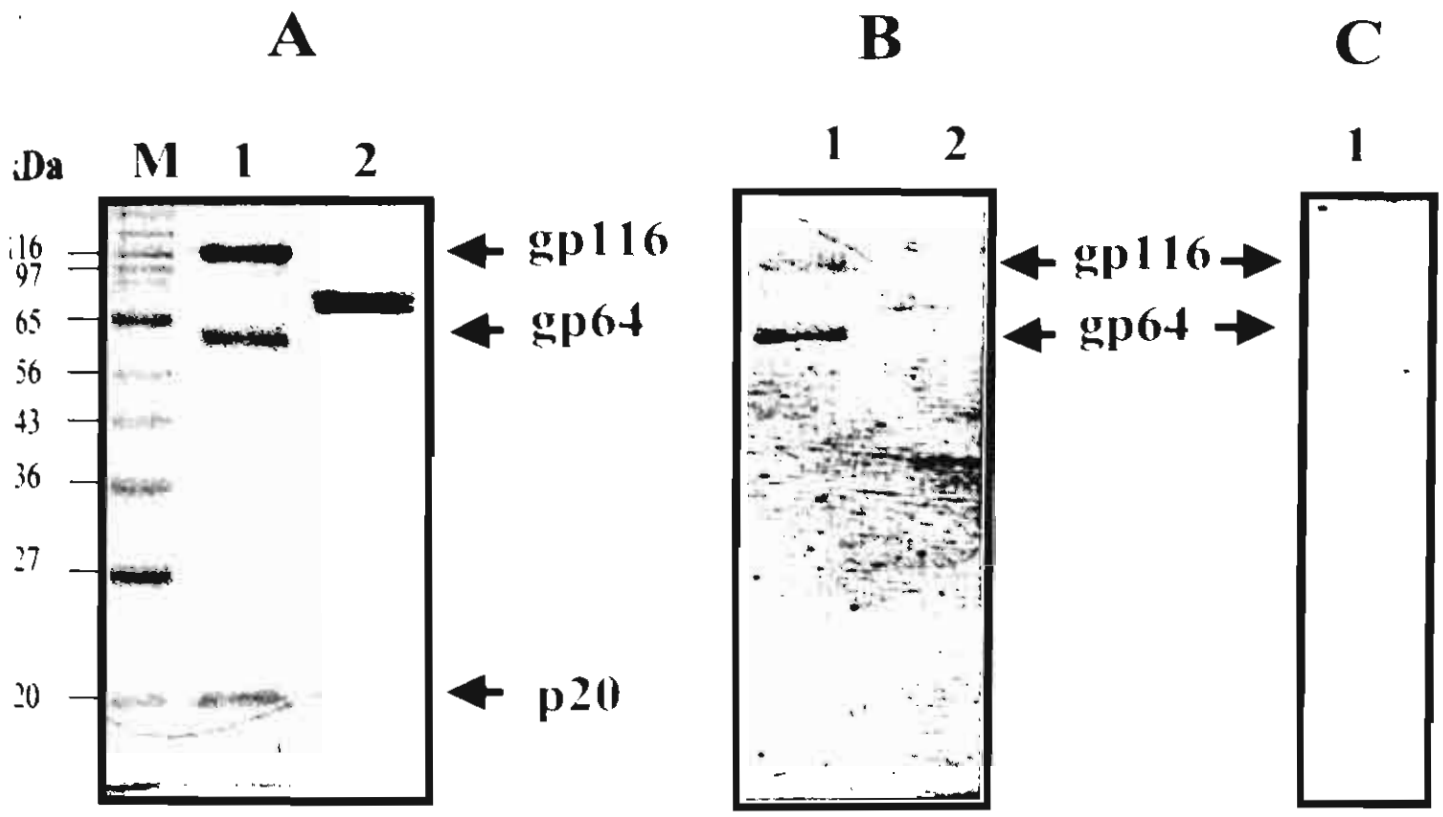
Figure 3. Nucleotide sequence and deduced amino acid sequence of YHV gp116 and gp64. Amino acid residues are numbered on the right. The amino acid residues identified by N-terminal sequencing of native gp116 and gp64 are in bold type. The predicted transmembrane helices are underlined. The potential *O*-linked glycosylation sites are shown in boxes. The proteolytic cleavages are shown by arrows. CCAAT box and TATA-like element are shown by double line underneath text.

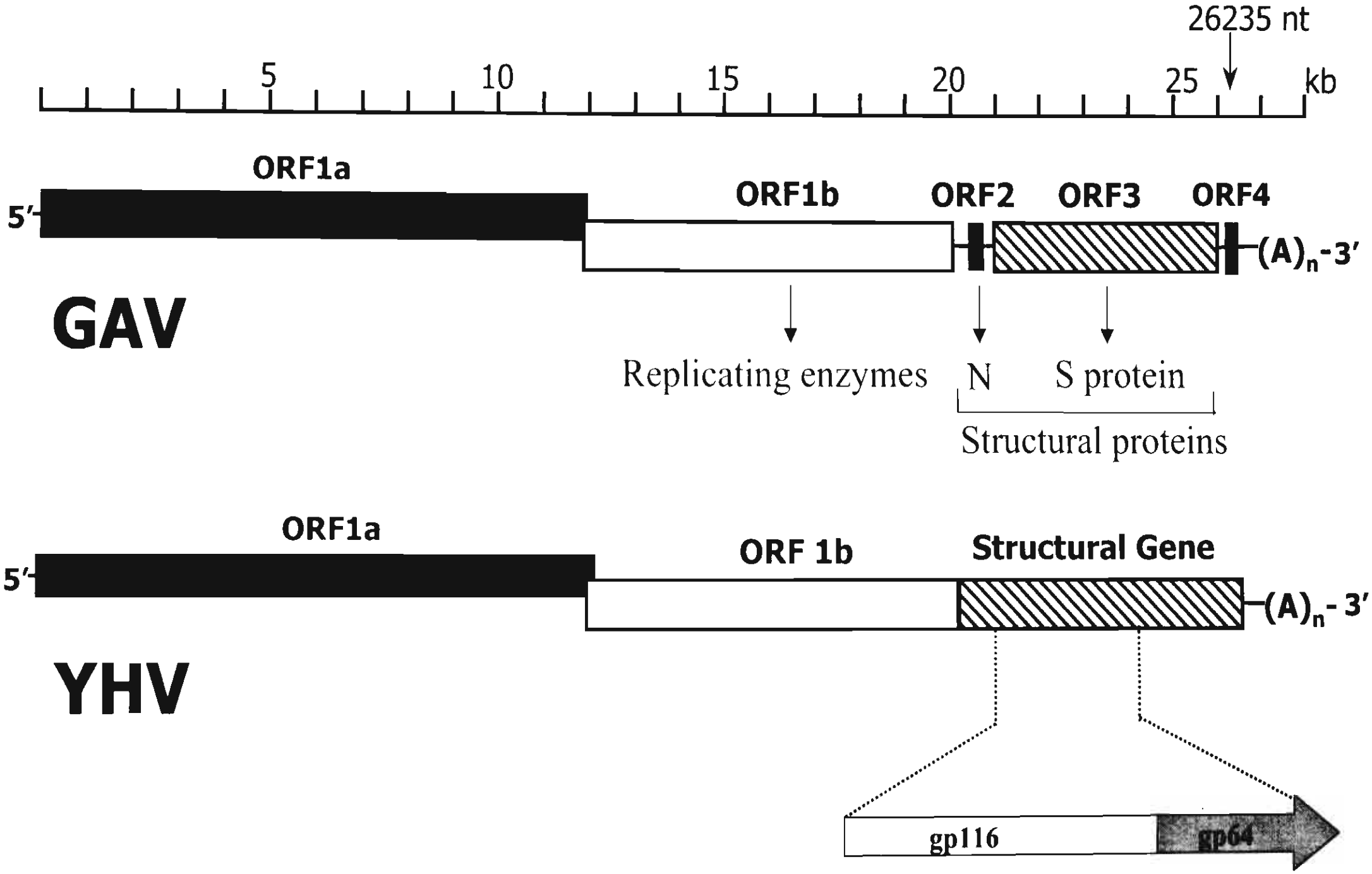
Figure 4. Predicted topology of YHV gp116 and gp64. Topology predictions were determined using TMHMM (Krogh *et al.*, 2001). Cylinders with the number represent the transmembrane domains while the solid lines represent ectodomain or intra-virion domains. Arrows represent the proteolytic cleavage sites.

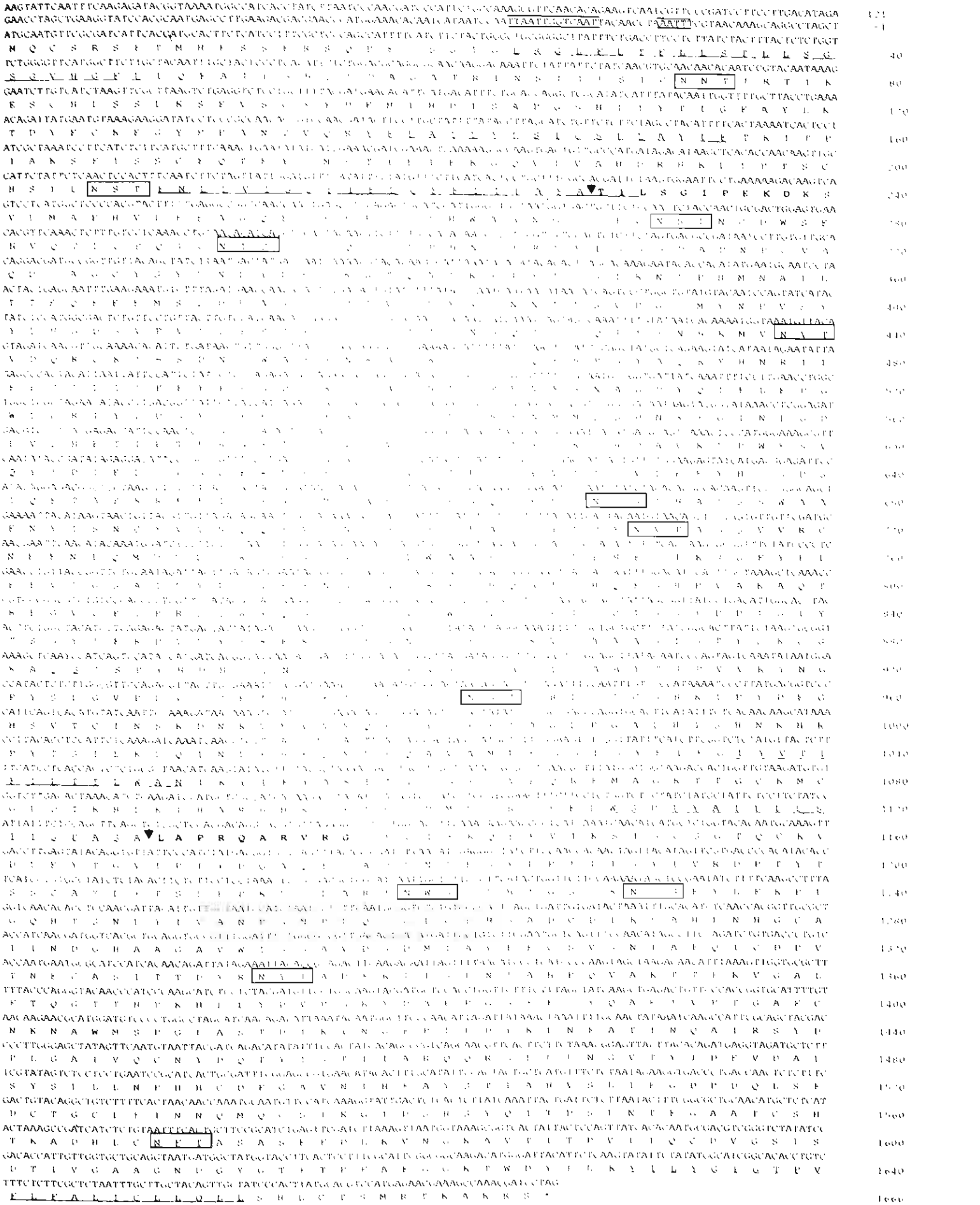
Figure 5. Comparison of protein encoded from ORF3 of YHV and GAV. Both sequences were aligned using Clustal W. (Thompson *et al.*, 1994). The conserved amino acids are shown as white text against black boxes. The residue number is also indicated.

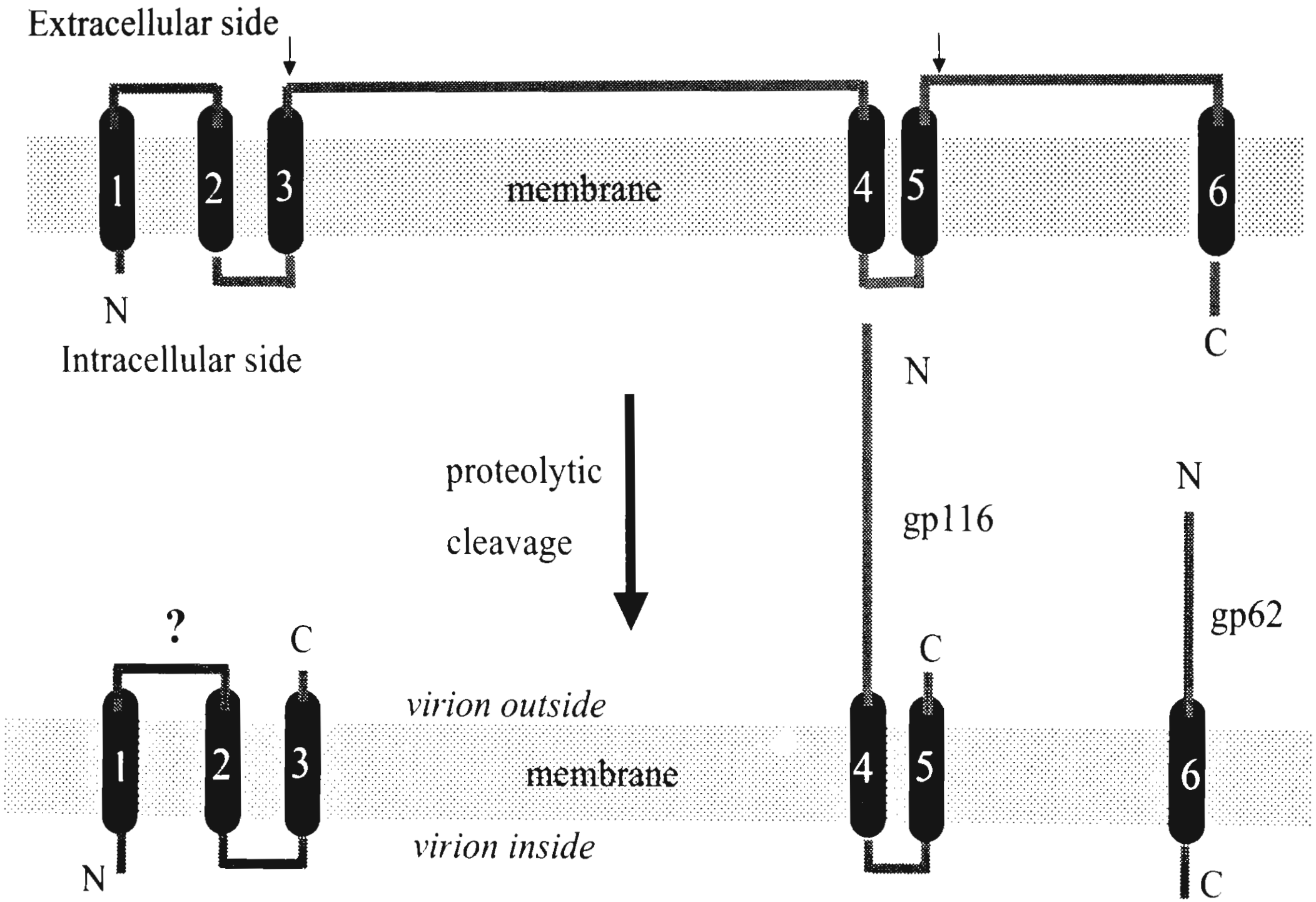
Figure 6. Expression of recombinant YHV gp64 with and without transmembrane in *E. coli*. (A) Schematic diagram of gp64 construct with or without transmembrane region. (B) Western blot of analysis of whole cell lysates of *E. coli* BL21(DE3) harbored

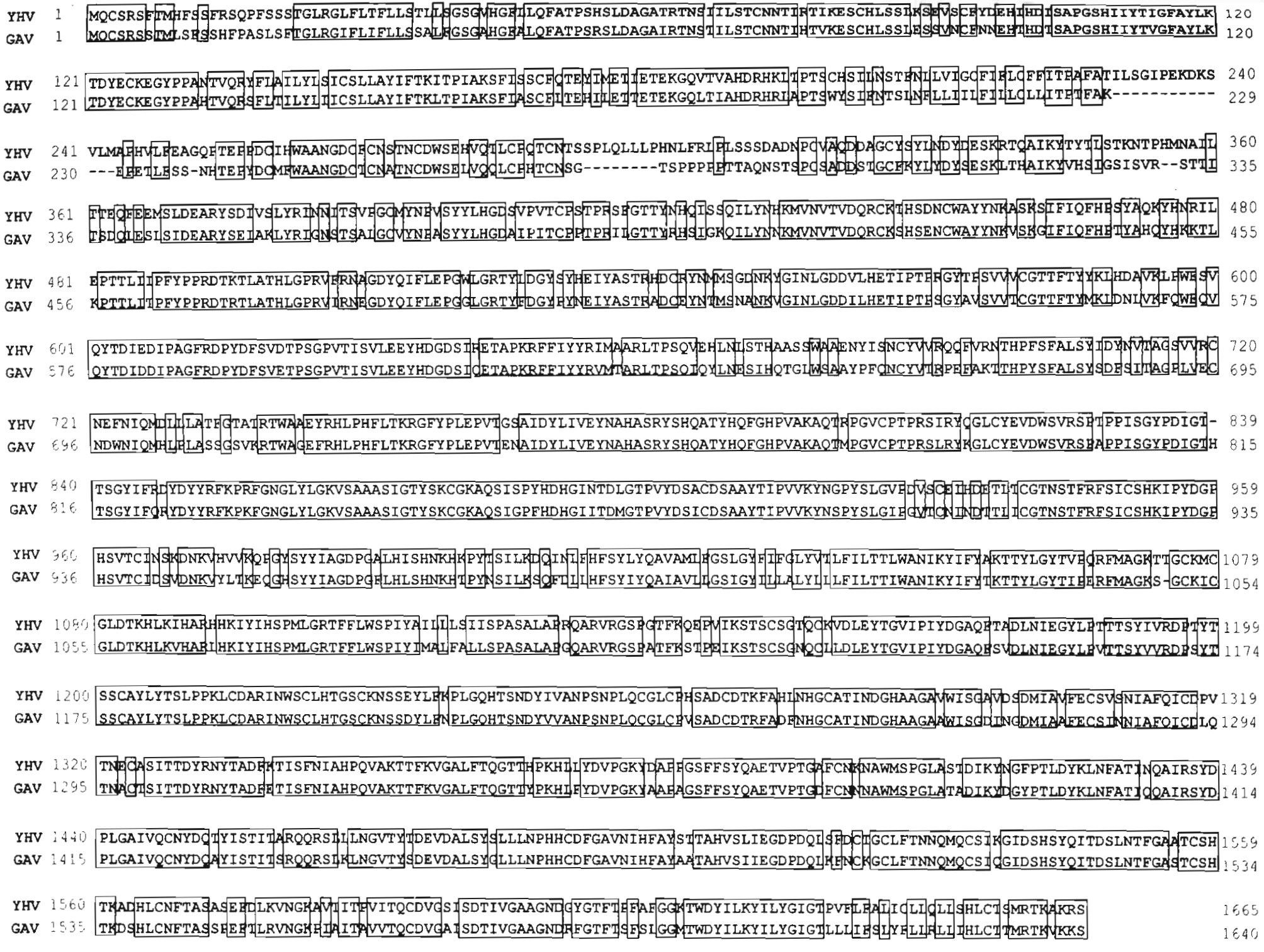
pET17b-gp64 induced for with IPTG. Immunoreactive proteins were detected using polyclonal antiserum raised against native gp64 in mice. M, molecular weight protein markers; 1, BL21(DE3) transformed with pET17b and induced for 3 h; 2-6, BL21(DE3) pLys transformed with pET17b-gp64, induced at 0, 1, 3, 5 and 7 h; 7, purified YHV. (C) BL21 (DE3) harbored pET17bgp64ΔTM (transmembrane deleted). M, molecular weight protein markers; 1, BL21(DE3) transformed with pET17b and induced for 3 h.; 2-4, BL21 (DE3) transformed with pET17bgp64ΔTM and induced for 1, 3 and 5 h. 5, purified YHV. Arrows on Western blot indicate the positions of the bands corresponding to the native gp64 and bacterially expressed gp64 with and without transmembrane region.


A
B
C
Da
M
1
2
1
2
1
16
97
65
56
43
36
27
20
gp116
gp64
p20
gp116
gp64

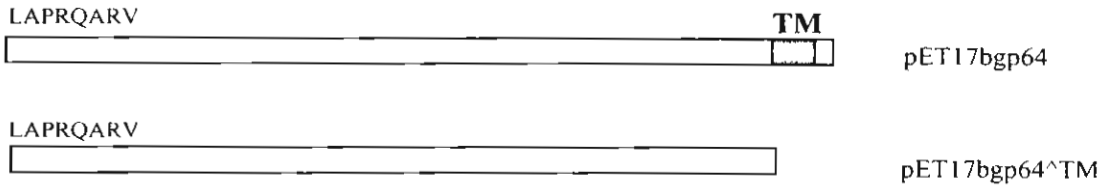

26235 nt
5
10
15
20
25
kb
ORF1a
5′
ORF1b
ORF2
ORF3
ORF4
(A)n-3′
GAV
Replicating enzymes
N
S protein
Structural proteins
ORF1a
5′
ORF 1b
Structural Gene
(A)n- 3′
YHV
gp116
gp64



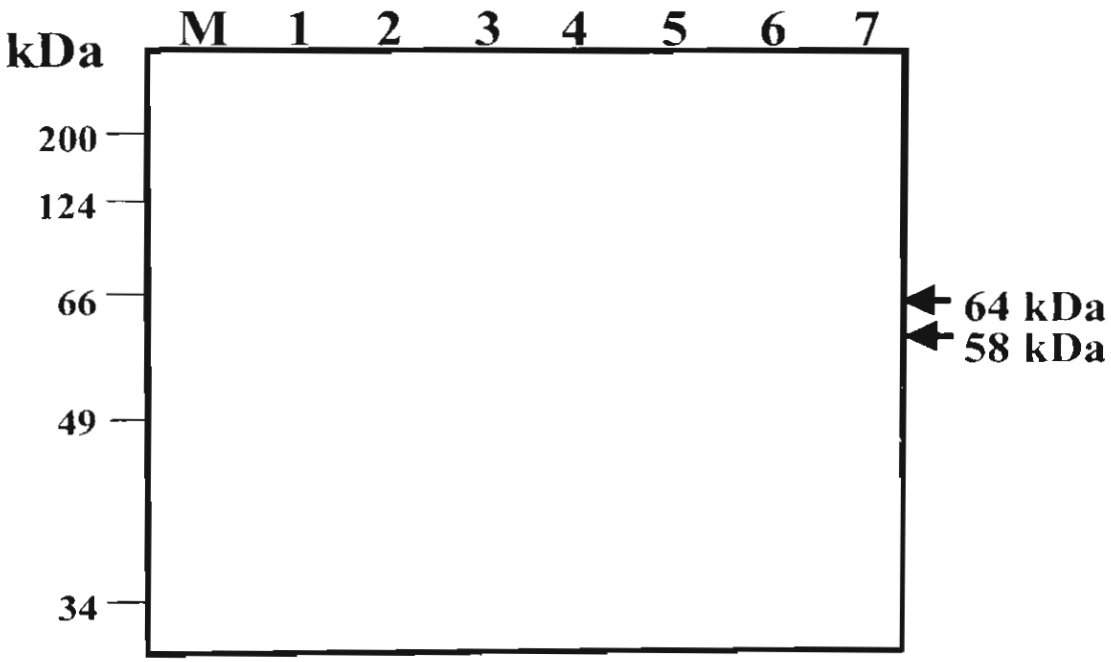

Extracellular side
1
2
3
membrane
4
5
6
N
C
Intracellular side
N
proteolytic
cleavage
gp116
N
gp62
?
C
C
virion outside
1
2
3
membrane
4
5
6
virion inside
N
C


YHV
GAV

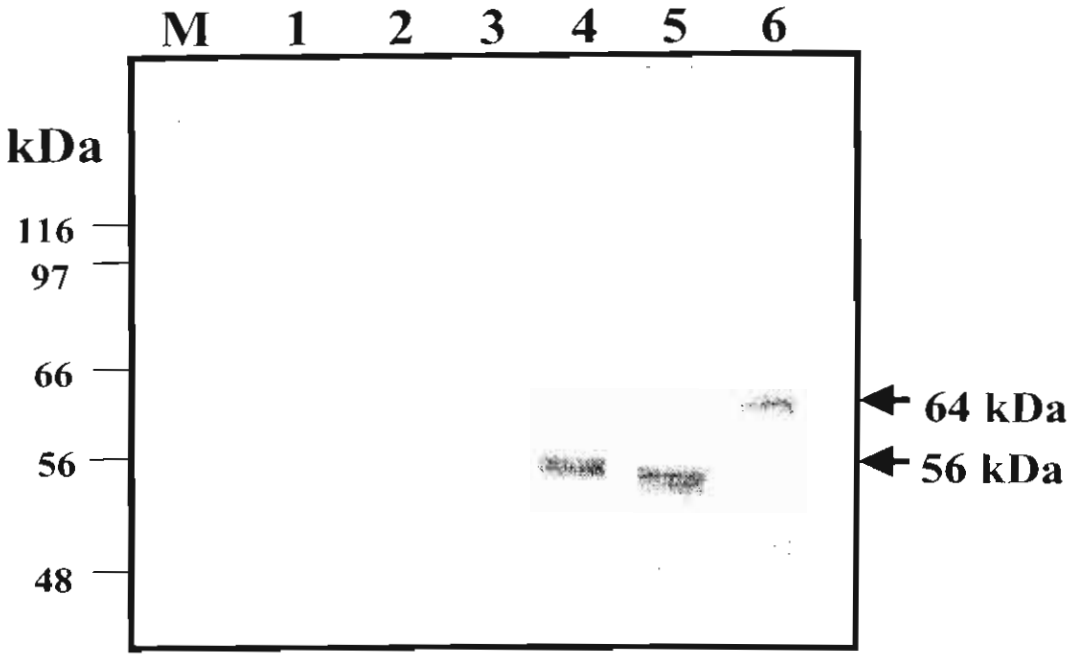
A


LAPRQARV
TM
pET17bgp64
LAPRQARV
pET17bgp64^TM

B


kDa
M 1 2 3 4 5 6 7
200
124
66
49
34
64 kDa
58 kDa

C


M 1 2 3 4 5 6
kDa
116
97
66
56
48
64 kDa
56 kDa